

ASSUMPTIONS IN EACH CLUSTERING TECHNIQUES -

1) K Means Clustering →

2 main assumptions are made by K means clustering:

- ① Clusters are spatially grouped or spherical. (grouped into cohort)
- ② Clusters are of similar size.

Suppose take one example,

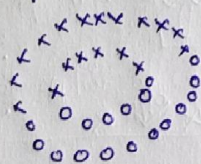


→ Two clusters right? One small ring (x) surrounded by large ring (o).

→ However, only one ring is a spherical cluster - the inner one. If we drew a circle around the outer cluster/ring it would have to encompass meaning surround and have or hold within the inner one.

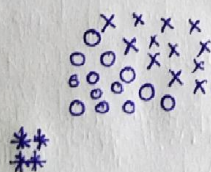
- K means lacks judgment in which similar size cluster in a dataset which is its assumption (small circle is small in size, large circle is large in size).

- When we apply K means output will be like

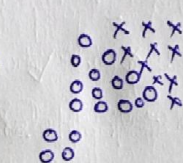


Instead of small circle as 'x' and big circle as 'o', K means will divide the population into similar size of each cluster.

- Another example,



K means
3 groups
originally
O x *

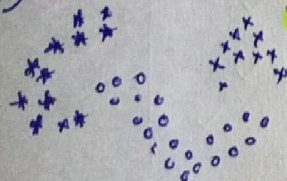


(Now 2 groups)
only - O, x

So instead of 3 clusters, K means reduces it to 2 because of intra cluster distances.

K means tries to produce tight clusters. In attempt to minimize the intra cluster distance between the points in the large cluster its "overdone" things & produce two clusters that have similar intra cluster distances.

2) DBScan → ① DBScan also make an assumption that clusters are dense regions in space separated by regions of lower density.



② A dense cluster is a region of densely connected, density of points in that region is minimum.

③ It can find clusters of arbitrary shapes based on density.

④ DBscan works well with the data which are severely affected by the presence of noise and outliers in the data.

Example of DBScan.

③ GMM Clustering -

- Non-Gaussian dataset \rightarrow GMM assumes an underlying Gaussian generative distribution.
- Uneven cluster size \rightarrow When clusters do not have even sizes, there is a high chance that smaller clusters get dominated by large one.
- Know the number of clusters (k) in advance.